



## UvA-DARE (Digital Academic Repository)

### Beyond binding: specialization without segregation

Scholte, H. Steven; de Haan, Edward H.F.

**DOI**

[10.1016/j.tics.2025.11.004](https://doi.org/10.1016/j.tics.2025.11.004)

**Publication date**

2026

**Document Version**

Final published version

**Published in**

Trends in Cognitive Sciences

**License**

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

**Citation for published version (APA):**

Scholte, H. S., & de Haan, E. H. F. (2026). Beyond binding: specialization without segregation. *Trends in Cognitive Sciences*, 30(3), 192-193.

<https://doi.org/10.1016/j.tics.2025.11.004>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Letter

Beyond binding:  
specialization without  
segregationH. Steven Scholte<sup>1,\*</sup> and  
Edward H.F. de Haan<sup>2,3,4</sup>From binding to natural  
composition

Roelfsema and Serre [1] contend that coherent object perception generally requires feature binding across segregated maps. We agree that selective control – attention, grouping, and recurrence – is essential when natural regularities are disrupted by occlusion, camouflage, isoluminance, or clutter. However, we observe specialization but no segregation: functional biases arise as local amplification within broadly overlapping, mixed-selectivity populations (see Box 1), not as isolated feature silos. As we argued [2], natural vision exploits the lawful structure of the world: co-occurring features such as curvature, material, and color–form statistics are encoded by overlapping populations from which many conjunctions are directly decodable in the feedforward pass. Composition is thus a property of representational geometry rather than evidence for a mandatory binding operation.

Binding, in this account, is a relational notion. In symbolic architectures, it refers to an explicit operation joining segregated stores, like combining symbols. In distributed population codes, it denotes the accessibility of conjunctions embedded in continuous, overlapping response spaces: comparable to retrieving coordinates from a vector field. The relevant question is therefore not whether the brain binds, but when explicit binding becomes necessary: only when statistical regularities fail to disambiguate relationships or when a task demands an explicit relational judgment.

Implicit composition and selective  
control

Under natural statistics, the feedforward sweep yields population geometries that mirror the structure of the visual world. Co-varying features – curvature spectra, material cues, color–form correlations – map to nearby trajectories, enabling downstream labeling (e.g., ‘hand’, ‘face’, ‘tree’) without explicit color × shape × motion operations [3].

We disagree with Roelfsema and Serre [1] on the default regime; they posit that segregated maps require binding, whereas overlapping, mixed-selectivity populations render many conjunctions directly accessible. We converge on the exceptional regime: when regularities fail or tasks demand explicit relations, binding via attention, grouping, and recurrence is necessary [1]. Control reweights or stabilizes existing geometry, improving separability in clutter or occlusion while offering little when feedforward codes suffice [4,5]; it refines representations rather than constructing tokens from scratch.

This distinction – implicit composition by default, explicit control on demand – is supported by converging neurophysiological and computational evidence. Networks trained on natural images develop hierarchical, mixed-selectivity codes where conjunctions emerge without enumerating combinations; late layers permit near-linear readout paralleling inferotemporal decodability [3,5,6]. Multi-task and world-modeling objectives yield graded overlap: related goals share subspaces, whereas unrelated ones diverge with depth [7,8]. Thus, specialized responses coexist with extensive representational overlap: specialization without segregation.

The same logic extends to behavior: in natural scenes, hands and faces are recognized almost automatically because their curvature, texture, and chromatic

statistics co-occur predictably. Only when relations must be resolved – whose hand, or a face through occlusion – do attention and recurrence become indispensable [4,5], explaining ventral sufficiency for routine recognition and fronto-parietal recruitment for relational, task-driven binding [2].

## Specialization ≠ segregation

Classical claims of segregation rested on comparisons confined to the extremes of the feature space – motion-only versus static form, saturated chromatic versus achromatic shape – rather than on dense, naturalistic sampling within the manifold. When responses are measured densely and analyzed multivariately, graded biases appear within broad overlap: neurons and voxels mix features across multiple dimensions, forming mixed-selectivity population vectors in a shared representational geometry rather than assigning each element to a single feature class [2,3].

Cortical mosaics exemplify specialization without isolation. In V4, color-biased globs and form-biased interglobs lie within continuous maps; both encode position and contour [9]. In MT/V5, strong motion tuning coexists with disparity and 3D structure, while MST and V3A extend to heading and position signals. Across millimeters, these domains form high-gain patches within a continuous sheet: specialization as amplification, not separation. Lesion and stimulation studies echo this pattern: selective deficits indicate reduced participation, not the absence of a feature store [10].

Population-geometry mapping across species reinforces this view. Macaque–human correspondences emerge from response geometry without predefined feature labels, revealing interleaved high-gain neighborhoods rather than segregated streams [11]. Task-optimized models reproduce this mosaic: related tasks occupy nearby manifolds; unrelated ones diverge smoothly [7]. Such graded, overlapping

### Box 1. Specialization ≠ segregation

We argue that specialization in visual cortex reflects local amplification within broadly shared codes, not sealed feature silos. When responses are sampled densely and analyzed multivariately, putative ‘modules’ resolve into overlapping, mixed-selectivity geometries. The following five points show how biased sampling, deficit-based inference, and microstimulation findings can overstate separation, and what to test instead.

(i) Sampling matters: comparisons confined to extremes (e.g., motion-only versus static form) exaggerate separation between features. Dense, naturalistic within-manifold sampling with multivariate analyses reveals graded biases within broad overlap [11]. Tuning should be interpreted as a vector pattern, not a categorical label.

(ii) Deficits ≠ modules: selective loss indicates high-gain impact, not exclusive coding. If other features were decodable from the same patch before damage, the code was overlapping. Defining specialization by deficit confuses causal influence with representational exclusivity.

(iii) Cortical mosaics: V4 and MT/V5 exemplify amplification in continuous sheets. V4 globs emphasize color yet encode form and position; interglobs favor form but retain color [9]. MT/V5 shows strong motion tuning plus disparity and 3D structure, with MST/V3A adding optic flow, heading, and position [10]. Natural-scene mapping reveals interleaved high-gain neighborhoods, not segregated streams [11].

(iv) Microstimulation: currents spread across local circuits, showing participation of the stimulated region but not the exclusivity of its code [12].

(v) Empirical signature: smooth decoding gradients: dense, naturalistic sampling should reveal co-decodable features in every ‘specialized’ patch, with graded – not absolute – effects, and smooth decoding gradients across high-gain patches and transition zones, yielding robustness to small misalignments and no sharp representational breaks.

organization reconciles the existence of local specializations with the distributed nature of cortical coding.

feature maps. See Box 1 for the empirical tests and predictions supporting specialization without segregation.

### Concluding remarks

Binding is not a unitary operation but a relationship between representational format and readout. Under natural statistics, overlapping, mixed-selectivity populations render many conjunctions directly accessible; when inputs or goals introduce ambiguity, selective control (attention, grouping, and recurrence) reweights and stabilizes the existing population geometry. In this view, specialization reflects local gain within a shared representational continuum, not a partitioning of the cortex into separate

### Declaration of interests

No interests are declared.

<sup>1</sup>Psychology Department, University of Amsterdam, 1001NK, Amsterdam, The Netherlands

<sup>2</sup>Donders Institute for Brain, Cognition and Behavior, Radboud University, 6525GD, Nijmegen, The Netherlands

<sup>3</sup>St Hugh's College, Oxford University, Oxford OX2 6LE, UK

<sup>4</sup>Psychology Department, Nottingham University, Nottingham NG7 2RD, UK

\*Correspondence:  
[h.s.scholte@uva.nl](mailto:h.s.scholte@uva.nl) (H.S. Scholte).  
<https://doi.org/10.1016/j.tics.2025.11.004>

© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

### References

1. Roelfsema, P.R. and Serre, T. (2026) Feature binding in biological and artificial vision. *Trends Cogn. Sci.* 30, 190–191
2. Scholte, H.S. and de Haan, E.H.F. (2025) Beyond binding: from modular to natural vision. *Trends Cogn. Sci.* 29, 505–515
3. DiCarlo, J.J. and Cox, D.D. (2007) Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341
4. Lamme, V.A.F. and Roelfsema, P.R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579
5. Kar, K. *et al.* (2019) Evidence that recurrent processing is required to classify objects in natural scenes. *Nat. Neurosci.* 22, 934–944
6. Bau, D. *et al.* (2017) Network dissection: quantifying interpretability of deep visual representations. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3319–3327
7. Scholte, H.S. *et al.* (2018) Visual pathways from the perspective of cost functions and multi-task deep neural networks. *Cortex* 98, 249–261
8. Tang, Y. *et al.* (2025) Diverse perceptual representations across visual pathways emerge from a single objective. *bioRxiv* Published online October 8, 2025. <https://doi.org/10.1101/2025.07.22.664908>
9. Conway, B.R. *et al.* (2007) Specialized color modules in macaque extrastriate cortex. *Neuron* 56, 560–573
10. Born, R.T. and Bradley, D.C. (2005) Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28, 157–189
11. Vinken, K. *et al.* (2025) Mapping macaque to human cortex with natural scene responses. *Proc. Natl. Acad. Sci. U. S. A.* 122, e2512619122
12. Tehovnik, E.J. *et al.* (2006) Direct and indirect activation of cortical neurons by electrical microstimulation. *J. Neurophysiol.* 96, 512–521